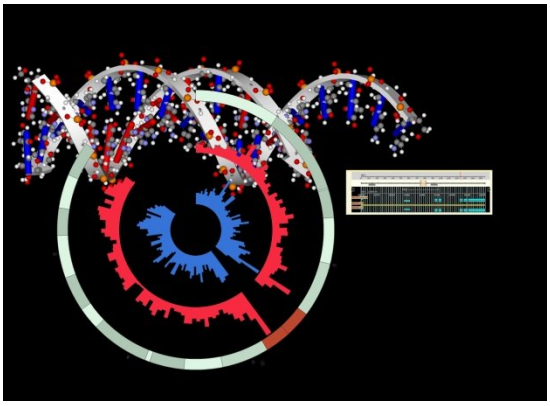
Accelrys NGS Collection for Pipeline Pilot Delivers Superior Performance on IBM iDataPlex and IBM SONAS – *Complete Human Genome Mapping in Hours, Not Days.*

Research teams using next generation sequencing (NGS) technologies are faced with the daunting challenge of supporting compute-intensive analysis methods against petabytes (PB) of data while simultaneously keeping pace with rapidly evolving algorithmic best practices. NGS users now can solve these challenges by deploying the Accelrys® Pipeline Pilot™ software platform on new systems offerings from IBM®.

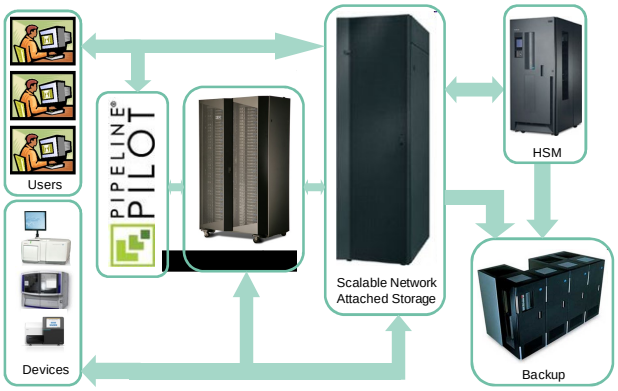


The Next Generation Sequencing (NGS) Collection for Accelrys Pipeline Pilot provides researchers with a versatile and agile platform to analyze and interpret the massive datasets generated by the most current DNA sequencing instruments. Using pre-built protocols, researchers can perform common computational workflows such as *de novo* sequencing, mapping to reference sequences, and variation detection; with the component collection, they can easily create other NGS protocols tailored to their needs. The NGS Collection supports native data formats from all of the major sequencing vendors allowing researchers to exploit the strengths of each sequencing platform and even combine results to augment analysis and interpretation.

Through use of a flexible, data management repository, Pipeline Pilot components access reference sequences, mapped reads, and genomic features. The data reader and writers support common formats, such as SAM, BAM, GFF3, or FASTQ, as appropriate, to optimize integration of industry-leading algorithms in this rapidly evolving domain. Because the NGS Collection is built on the Pipeline Pilot informatics platform, third-party applications can be integrated and new algorithms can be easily integrated into existing data pipelines. This enables researchers to become significantly more productive.

The IBM / Accelrys Advantage for NGS

Faster computing not only saves money (each day's delay in bringing a drug to market costs millions of dollars) but also takes researchers closer to the goal of $1000 for a human genome that medical researchers believe will make next generation sequencing affordable and transform preventive and prescriptive medicine.

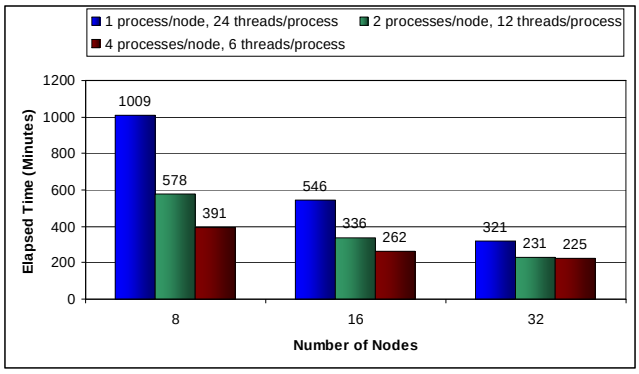


Pipeline Pilot NGS System Architecture

A compute architecture using IBM® System x® iDataPlex™ systems and IBM SONAS® (Scale Out Network Attached Storage) based on IBM's flagship General Parallel File System (GPFS™) delivers outstanding performance for such demanding requirements.

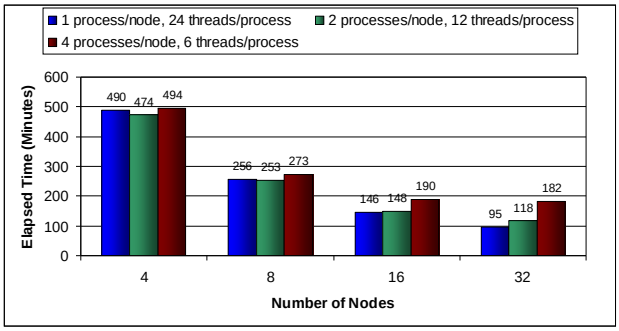
For example, it takes about two to three days to complete human genome mapping with typical 30x coverage with the widely-used BWA algorithm on a single node. With Pipeline Pilot BWA Mapper in an IBM compute architecture, researchers have achieved:

- Complete mapping in <4 hr with 32 iDataPlex system nodes with GPFS
- Complete mapping in about 4.5 hr with 16 iDataPlex system nodes with GPFS
- Complete mapping in 6 hr with 8 iDataPlex system nodes with GPFS



Similarly, it takes a few days to complete human genome mapping with typical 30x coverage using the open source algorithm Bowtie on a single node. Running the Bowtie mapping pipeline on Pipeline Pilot in an IBM compute architecture, researchers have achieved:

- Complete mapping in ~ 1.5 hr with 32 iDataPlex system nodes with GPFS
- Complete mapping in ~ 2.5hr with 16 iDataPlex system nodes with GPFS
- Complete mapping in ~ 4 hr with 8 iDataPlex system nodes with GPFS



The IBM iDataPlex Solution

IBM System x™ iDataPlex solution can help researchers:

- Double the number of servers you can run in a single rack for better space utilization
- Use up to 40% less energy while increasing data center computing power five times
- Reduce air conditioning significantly.

IBM SONAS

IBM SONAS leverages mature technology from IBM's high performance computing (HPC) experience. Based on IBM's flagship General Parallel File System (GPFS), SONAS is an easy-to-install, turnkey, modular, scale out Network Attached Storage (NAS) solution that provides the performance, clustered scalability, high availability (HA) and functionality that are essential to meeting strategic PB Age and cloud storage requirements.

More information

Accelrys NGS Collection for Pipeline Pilot:
http://accelrys.com/products/pipeline-pilot/component-collections/next-generation-sequencing.html

IBM System x iDataPlex solutions:
www.**ibm.com**/systems/x/hardware/idataplex/

IBM SONAS
http://www-03.ibm.com/systems/storage/network/sonas/

IBM Corporation
Route 100 Somers, NY 10589 U.S.A.
Produced in the United States of America
April 2011

The benchmarks and values shown here were derived using particular, well configured, development-level computer systems. All benchmark values are provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Actual system performance may vary and is dependent upon many factors including system hardware configuration and software design and configuration. Buyers should consult other sources of information to evaluate the performance of systems they are considering buying and should consider conducting applications oriented testing. Test case description: In this case we sought to study the effect of increasing the number of nodes and threads on the iDataPlex solution for two Next Generation Sequencing (NGS) applications, BWA Mapper and Bowtie Mapper, both part of Accelrys Pipeline Pilot 8.0.

This publication was developed for products and/or services offered in the United States. IBM may not offer the products, features, or services discussed in this publication in other countries. The information may be subject to change without notice. Consult your local IBM business contact for information on the products, features and services available in your area. This publication is based on information provided by Accelrys Software, Inc. Many factors contributed to the results and benefits described. IBM does not guarantee the same or similar results for all clients. IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, IBM warranty terms apply. All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only.